

Andrej Perdih, Nina Ledinek

ZRC SAZU, Fran Ramovš Institute of the Slovenian Language, Ljubljana

Multi-word Lexical Units in General Monolingual Explanatory Dictionaries of Slavic languages

Članek obravnava tipologijo in makro- oz. mikrostrukturno umeščanje večbesednih leksikalnih enot v petih enojezičnih splošnih razlagalnih slovarjih slovanskih jezikov, in sicer slovenskem, hrvaškem, slovaškem, poljskem in ruskem. Večbesedne leksikalne enote so v teh slovarjih obravnavane kot iztočnice in podiztočnice, med slovarskimi zgledi, v okviru posameznih slovarskih pomenov in v posebnih mikrostrukturnih razdelkih, npr. frazeološkem, paremiološkem, terminološkem. Razlike v obravnavi večbesednih leksikalnih enot v analiziranih slovarjih so posledica različnih jezikovnoteoretičnih pristopov, strukturiranja slovarskih podatkov na ravni slovarja in slovarske baze in tudi medija, za uporabo v katerem je bil slovar primarno oblikovan. Članek se osredotoča tudi na večbesedne leksikalne enote v eSSKJ: tretji izdaji Slovarja slovenskega knjižnega jezika, v najnovejšem splošnem enojezičnem razlagalnem slovarju slovenščine, v katerem so te enote obravnavane kot relativno samostojne enote, opisane na podoben način kot enobesedne leksikalne enote.

The article analyses the typology and the macro- and microstructure positioning of multi-word lexical units (hereinafter: MLUs) in general monolingual explanatory dictionaries of five Slavic languages: Slovenian, Croatian, Slovak, Polish and Russian. MLUs in these dictionaries are included as headwords, as subheadwords, as examples, as independent senses, and as units in various types of MLU sections (e.g. phraseological, terminological, or paremiological sections). Differences regarding the positioning and representation of MLUs in these dictionaries arise from different theoretical backgrounds and data structuring as well as the medium in which a dictionary was first published. With regard to the findings about the typology and positioning of MLUs in these dictionaries, the article also focuses on the treatment of MLUs in eSSKJ: The Dictionary of the Slovenian Standard Language, 3rd edition, the new monolingual general explanatory dictionary of the Slovenian language. In this dictionary, MLUs are treated in a similar way to single-word lexical units and are given relative autonomy in the dictionary structure.

0 Introduction

Multi-word lexical units (hereinafter: MLUs) represent a large part of the lexicon of a language. The different types of MLUs (phraseological MLUs, terminological

MLUs, and lexicalized MLUs – i.e. non-phraseological non-terminological MLUs) can be treated differently even in dictionaries of the same type and placed in several microstructural positions as well as on the macrostructural level. MLUs can be presented as headwords and subheadwords, or may be included among the examples within a specific dictionary sense. Additionally, MLUs may be placed in the phraseological section, paremiological section (for proverbs), terminological section, and lexicalized-MLUs section. These sections (and subsequently their MLUs) may be treated within a dictionary sense if they are semantically related to it. Alternatively, these sections may be placed separately from dictionary senses. MLUs can be afforded a relative autonomy in which case they are dealt with in ways similar to single-word units so that labels, usage information, definitions, examples, etc. are added to describe MLUs.

It seems that there is no consensus among lexicographers as to where in the macro- or microstructure of the dictionary entry the MLUs of different types should be treated. According to Atkins & Rundell (2008: 394) and Vrbinc (2011), the treatment of MLUs differs greatly among English dictionaries. Kržišnik (2009) conducted research on MLUs in English, German, Russian, Croatian, Slovak, and Slovenian general and pedagogical monolingual explanatory dictionaries, concluding that among the analysed dictionaries some English dictionaries place MLUs in the headword position. In general, however, MLUs are mostly treated in separate microstructural sections, usually in a phraseological section or in a non-phraseological section. Oppentocht & Schutz (2003: 219) advocate a structured collection of types of lexical entities where a distinction between single words, fixed phrases (collocations and idioms), and free text is the basic organizing principle. Vrbinc (2011) recommends that English learners' dictionaries and the bilingual English-Slovenian dictionary include MLUs in two places (as headwords and as idioms in a separate idioms section in a single-word entry) since "it is next to impossible to predict whether users will look up a MLU as a headword or will simply look up one of the constituent elements of such a lexical item (but which one?)",¹ according to the study on dictionary use of Slovenian learners of English presented in Vrbinc & Vrbinc (2004).

¹ In the context of improving search possibilities, Lew (2012) describes various approaches to searching phraseological units, which depend on the users' search strategies and the technical possibilities of electronic-dictionary search engines. Cf. also the search strategies available on the Fran dictionary portal (Ahačič et. al 2015) and the above-mentioned proposal by Vrbinc.

1 Typology and positioning of MLUs in general monolingual explanatory dictionaries of Slavic languages

In this paper, we treat MLUs as combinations of two or more lexemes, where a combination becomes a new lexical unit carrying a meaning of its own. Excluded are multi-word citations from foreign languages that in the target language represent a single lexical unit although written separately (*a cappella*, *ad hoc*), reflexive verbs (*bati se*), and other types where constituents do not exist as single-word lexemes (*čira čara*) (Gliha Komac et al. 2015: 7).

Our research on the typology and positioning of MLUs in dictionaries of Slavic languages was conducted on the following dictionaries:²

- *Dictionary of the Slovenian Standard Language / Slovar slovenskega knjižnega jezika* (1st edition 1970–1991, 2nd edition 2014)
- *Great Dictionary of the Croatian Language / Veliki rječnik hrvatskoga jezika* (2003)
- *Dictionary of Contemporary Slovak / Slovník súčasného slovenského jazyka* (2006–)
- *Great Dictionary of Polish / Wielki słownik języka polskiego PAN* (2008–)
- *Explanatory Dictionary of Russian Language Including Explanations of Word Origin / Tolkovyj slovar' russkogo jazyka s vklyucheniem svedenij o proishozhdenii slov / Толковый словарь русского языка с включением сведений о происхождении слов* (2008)

Among these dictionaries, only the Great Dictionary of Polish is designed for use in an online digital environment. Due to the new possibilities of visualization, data organization, hyperlinking, different space constraints, etc. in digital environment, we can expect different lexicographic treatment of MLUs in digitally-born dictionaries as opposed to print dictionaries.

1.1 Dictionary of the Slovenian Standard Language, first and second edition

The dictionary (hereinafter: SSKJ) was compiled at ZRC SAZU, the Fran Ramovš Institute of the Slovenian Language. It was first published in 1970–1991 and updated in its second edition in 2014 (Perdih & Snoj 2015). No fundamental conceptual changes were introduced in the second edition and the treatment of MLUs also remained unchanged. The dictionary represents the only (finished)

² A part of the research on dictionaries of Slavic languages was presented in Perdih (2016). Regarding treatment of MLUs in dialectal and historical dictionaries, and the Slovenian dictionary of synonyms, see for example Žugić (2016), Legan Ravnikar (2017) and Snoj (2019: 185-193).

general monolingual explanatory dictionary of the Slovenian language and contains almost 100,000 entries. Both editions are freely available at the Fran dictionary portal (www.fran.si).³ MLUs in this dictionary are presented in several microstructural positions:⁴

- examples,
 - senses,
 - phraseological sections,
 - and terminological sections.
- (1) **abecéda** [...] **1.** [...] ♦ elektr. Morsejeva abeceda iz pik in črt sestavljeni znaki za br-zojavni prenos sporočil **2.** ekspr. začetno, osnovno znanje [...]
 - (2) **absolúten** [...] **1.** [...] **2.** [...] **3.** [...] ♦ filoz. absolutni duh
 - (3) **hidrát** [...] kem. [...] **2.** v zvezi ogljikov hidrat organska spojina ogljika z vodikom in kisikom, ki sta v takem razmerju kot v vodi
 - (4) **hišen** [...] / hišni pes pes, ki se goji v stanovanju, zlasti za družbo, zabavo; hišni red pravila o pravicah in dolžnostih stanovalcev, navadno v večstanovanjski hiši [...] ● iron. v stanovanju je našel hišnega prijatelja ženinega ljubimca; ekspr. v razpredelnico je napisal same hišne številke izmišljene, neprave ♦ zool. hišni pajek pajek, ki prede lijakaste mreže po kotih, *Tagenaria domestica*

MLUs are included in four different microstructural elements. MLUs (4) *hišni pes* and *hišni red* are presented among examples, *hišni prijatelj* and *hišna številka* are placed in the phraseological section (starting with the character ●). The terminological unit *hišni pajek* is placed in the terminological section (starting with the character ♦). The terminological MLU (3) *ogljikov hidrat* is presented as a separate sense of the word *hidrat*.

While representation among examples can occur only within a sense (4), both the terminological and phraseological sections can hierarchically either be included in a specific sense as represented by (1) ♦ *Morsejeva abeceda* or independently, i.e. after the last sense as shown in (2) ♦ *absolutni duh*. The distinction is based on the level of semantic shift between a sense and an MLU. Besides phraseological and paremiological units, the phraseological section (as it is called in the front matter of the dictionary) also contains various MLUs that do not fit typologically into clearly defined groups due to their divergence from other senses while often being too infrequent to be presented in a sense of their own (Silvester 1978: 37).

³ In the search result list, entries from the second edition are presented to the user, while entries from the first edition are available in the archive.

⁴ In entries (or their parts) that are used as examples in this paper, MLUs are underlined by the authors of the paper in order to easily find MLUs since the visual presentation differs from dictionary to dictionary.

MLUs included among examples are delimited from preceding examples by a slash sign “/” and, when necessary, accompanied by labels and definitions. However, in many cases they are not presented explicitly as standalone MLUs, but rather in contextual usage as part of an example. This also holds for the units in the phraseological section. When MLUs such as *ogljikov hidrat* are represented on the sense level (3), usage information *v zvezi* “in phrase”, *navadno v zvezi* “usually in phrase” (when a MLU is written explicitly, as a whole), *v zvezi s/z* “in combination with”, and *navadno v zvezi s/z* “usually in combination with” (when only the additional word is presented) is added.

1.2 Great Dictionary of the Croatian Language

This dictionary of 70,000 entries was published in 2004 by Novi Liber. It represents the latest general monolingual explanatory dictionary of Croatian language.⁵ MLUs are represented in the following positions:

- phraseological sections,
 - non-phraseological sections (terminological and non-terminological),
 - and examples.
- (5) **kùć|a** ž 1. zgrada koja ima zidove i krov i služi za stanovanje [*kamena ~a* kuća od kamena; *zidana ~a* kuća od opeke; (za razliku od drvene ili kuće od čerpiča)]; hiža [...] Δ **Božja ~a** *pren.* crkva; **javna ~a** bordel, mjesto gdje se novcem kupuje žensko društvo [...] □ **dogovor ~u gradi** dogovorom se postiže napredak; **(biti) kao kod svoje ~e** dobro se snalazit [...]

The entry (5) *kuća* contains all three locations where MLUs can be placed in the dictionary. While MLUs *kamena kuća* and *zidana kuća* are placed among examples and definitions are added, the non-phraseological section (starting with the character Δ) contains MLUs *Božja kuća* and *javna kuća*, whereas the phraseological section (starting with the character □) contains phraseological units *dogovor kuću gradi* and *(biti) kao kod svoje kuće*.

1.3 Dictionary of Contemporary Slovak

The dictionary is being compiled at the Ľ. Štúr Institute of Linguistics, Slovak Academy of Sciences. The first volume was published in 2006. The final scope of the dictionary is expected to reach 220,000 entries. So far three volumes have been published, containing entries from A to N published in print and web form (<http://slovník.juls.savba.sk/>).

⁵ The dictionary is based on the 3rd edition of the Dictionary of the Croatian Language (1998).

Semantically non-transparent MLUs in this dictionary can be located in the following locations:

- examples,
 - senses,
 - phraseological sections,
 - paremiological sections,
 - lexicalized-units sections,
 - terminological-units section,
 - and subheadwords.
- (6) **bielo** prisl. [...] do biela ↗ dobiela [...]
- (7) **dobiela**, pís. i do biela prisl. [...]
- (8) **dostat'** [...] **12.** ► ako formálne sloveso tvorí s pripojeným podstatným menom lexikalizované spojenie: d. doživotie byť odsúdený na doživotný trest; d. guráž odhodlať sa, osmeliť sa
- (9) **dupkom** prisl. **1.** iba v spojeniach vlasý mu stoja, vstávajú, stáli dupkom ► rovno dohora, priamo, ↗ i fraz. **2.** iba v spojení dupkom hľadieť ► meravo [...]
- (10) **dôvod** -du pl. N -dy m. [...]
 ■ **z dôvodu** predl. s G kniž. [...]
- (11) **charakteristika** [-t-t] -ky -tik ž. ⟨gr.⟩ **1.** ► vystihnutie podstatných, typických vlastností niekoho, niečoho; literárny žáner založený na vystihnutí podstatných, typických vlastností niekoho, niečoho [...] □ mat. *charakteristika logaritmu čísla* celá časť tohto logaritmu
- (12) **chlapský** [...] hovor. chlapská pijatika veľmi silná [...] pren. chýba tu chlapská ruka chlap, muž, otec ako zábezpeka rodiny, posila v domácnosti, pri výchove detí
- (13) **chvost** [...] ■ fraz. držať slovo ako pes chvost nevedieť dodržať sľub
- (14) **chvost** [...] ■ porem. bol by chytil zajaca, ale mal krátky chvost vyhovára sa
- (15) **chlípať** [...] pren. Ťažko sa chlípe s cudzou lyžicou. [J. Čajak ml.] ťažko sa žije mimo domova, v cudzine
- (16) **choroba** -by -rôb ž. **1.** ► fyzická al. psychická porucha zdravia, narušený zdravotný stav [...] □ horská choroba dýchacie problémy prejavujúce sa nedostatočným prísunom kyslíka vo veľkej nadmorskej výške [...]

In the first volume of the dictionary, MLUs are positioned in four different MLU sections. The phraseological section starts with the character ■ followed by the fraz. 'phraseology' abbreviation (13) (držať slovo ako pes chvost). The paremiological section starts with the character ■ followed by the porem. 'paremiology' abbreviation (14) (bol by chytil zajaca, ale mal krátky chvost). The lexicalized-units section (containing non-terminological non-phraseological units) is introduced by an orange square sign (□). In a lexicalized-units section, the headword is semantically shifted as in the case of (16) horská choroba. Multi-word proper names are also included in this type of sections. In the second volume of the dictionary the terminological section

(introduced by a black square sign ◻) was established, thus graphically differentiating multiword terminological units from non-terminological units, hierarchically positioned under the same sense as is the case of (11) *charakteristika logaritmu čísla*. Terminological proper names are treated in the same manner.

When the headword (or one of its senses) is part of one or a few MLUs and therefore has limited collocates, it is described on the sense level and usage information such as *iba v spojení/spojeniach* “only in combination”, *obyč. v spojení/spojeniach* “usually in combination”, *vo fraz. spojení/spojeniach* “in phraseological combination(s)” is added (SSSJ 26–27) as shown in (9) (*vlasý mu stoja, vstávajú, stáli dupkom; dupkom hladiet*). Adjectives used only as a part of MLUs, however, are not treated as dictionary senses and only MLUs (with definitions) are presented. When a headword has, in a certain sense, acquired figurative meaning, it is presented among examples, and short definitions are added to the examples. From the user’s point of view it is not always clear whether such examples represent the use of an MLU or just a semantic shift of a single-word unit. This is especially true for non-typical usage that is included in the dictionary.

Examples with short definitions are used when the headword’s meaning is shifted in certain MLUs. This approach is used on citations representing lesser-used meanings as shown in (15) (*Ťažko sa chlípe s cudzou lyžicou*) or semantically creative usages of a headword (12) (*chýba tu chlapská ruka*). Verbs are treated slightly differently, since all such types are merged into a single sense (8) (*dostať doživotie, dostať guráž*).

Secondary multi-word prepositions composed of a preposition and a fixed form of a lexical word are treated in full as subheadwords (10) (*z dôvodu*). Secondary multi-word adverbs (6) (*do biela*) are included as subheadwords with cross-references to their own entries where they are spelled as a single word or as separate words (7) (*dobiela, do biela*).

1.4 Great Dictionary of Polish

The Great Dictionary of Polish is an ongoing project coordinated by the Institute of Polish Language at the Polish Academy of Sciences. The number of published entries was expected to reach 50,000 entries in 2018, excluding MLUs (Žmigrodzki 2014: 40, Žmigrodzki 2018: 14); however, at the time of writing, the number of single-word entries reached 70,496, while the number of MLUs rose to 16,825.⁶ The dictionary is, however, expected to reach a larger volume in the following years. It is available online at www.wjssp.pl.

⁶ The authors kindly thank the editor-in-chief, Piotr Žmigrodzki, for the information on the current state of the dictionary (6 July, 2019).

Beside single-word lexemes, the dictionary contains MLUs that are treated separately as a special type of entries that are included in two lists:

- list of proverbs,
- and MLUs list.

ręka

Hasło ma wiele znaczeń,
wybierz to, które Cię interesuje

1.a kończyna

1.b dłoń

2. w piłce nożnej

3. mistrza

4. władza

5. małżeństwo

FRAZELOGIZMY: ✕

atut w ręku
atut w ręce
 bez kogoś/czegoś jak bez ręki
 brudne ręce

PRZYSŁOWIA: ✕

Uderz w stół, a nożyce się odezwą
Uderz ręką w stół, a nożyce się odezwą

Figure 1: Entry *ręka* containing a list of MLUs (*frazelogizmy*) and proverbs (*przysłowia*). The lists are shortened for this presentation.

atut w ręku

rzadziej *atut w ręce*

decydujący, przekonujący argument pozwalający
na uzyskanie przewagi i panowanie nad sytuacją

POKAŻ WSZYSTKO

Lubię to! 0

Data ostatniej modyfikacji: 04.09.2013

Warianty
Kwalifikacja tematyczna
Połączenia
Cytaty
Odmiana
Pochodzenie

Figure 2: MLU *atut w ręku* as a standalone special entry type.

Paremiological units (*przysłowia*) are separated from other types of MLUs (*frazeologizmy*). Both types are treated as independent entries; therefore, they are not really nested, but rather listed and cross-referenced (linked) within single-word entries. Both MLU groups contain nearly the same types of information as single-word lexemes, such as variants, labels, definitions, ontology, lexical relations (synonyms, antonyms, etc.), collocations, citations, morphological and syntactical information. Information about the origin of MLUs is given for well-documented borrowings from other languages. Proverbs can also contain text modifications.

The MLUs are treated as separate units and are not semantically referenced to specific senses of single-word lexemes (Żmigrodzki 2015: 65-68). Technically, they can be found either in lists within single-word entries or as search results. In both cases they are hyperlinked to their standalone entries.

1.5 Explanatory Dictionary of Russian Language Including Explanations of Word Origin

This dictionary, written by Shvedova, contains 82,000 lexical units, consisting of one-word lexical units and phraseological MLUs. It is based on Explanatory Dictionary of Russian Language, 4th edition, by Ozhegov and Shvedova that has a long tradition in Russian lexicography. MLUs are represented in the following locations:

- examples,
 - phraseological sections,
 - and after headwords.
- (17) **БАКЛУШИ:** бить баклуши (разг.) – бездельничать [*первонач.* делать несложное, лёгкое дело – разбивать полено на баклуши, т. е. чурки для выделки мелких изделий].
- (18) **ОПРЕДЕЛЕНИЕ**², -я, *ср.* В грамматике: второстепенный член предложения — атрибут, обычно выражаемый согласуемой формой и обозначающий собственно признак, свойство или качество. Согласованное о. (выражаемое прилагательным).
- (19) **ХАЛЯВА**, -ы *ж.* (прост, груб.). [...] ♦ на халяву – 1) кое как, небрежно *Что ни сделает, всё на халяву*; 2) бесплатно, задаром. *Досталось на халяву.*
- (20) **ХАРАКТЕР**, -а, *м.* 1. Совокупность психических, духовных свойств человека, обнаруживающихся в его поведении. *Сильный, волевой, твёрдый, смиренный х. Выдержать х.* (сохранить твёрдость, не уступить в чем н.).

MLUs can be placed in several positions in the microstructure of a dictionary entry. As has been shown for other dictionaries, MLUs can be placed among examples also in this dictionary (20) (*выдержать характер*). Terminological

MLUs are treated similarly, since they are included in senses that are introduced by terminological labels and defined (18) (*согласованное определение*). MLUs in the phraseological section (introduced by the character ♦) are defined and examples are added to represent their usage (19) (*на халяву*). Additionally, words that appear only as parts of a specific MLU, are presented on a headword level with the MLU directly following the headword and being treated as other MLUs (17) (*бить баклуши*).

1.6 Discussion

Similarly to previous research on different dictionaries (as mentioned above) also the presented dictionaries of Slavic languages show high variety in placing and distinguishing various MLU types.

	SSKJ	VRHJ	SSSJ	WSJP	TSRYA
headword	–	–	–	+/_*	–
subheadword	–	–	+	–	–
MLU sections	Ph, T	Ph, T	L, Ph, Pa, T	Phl, Pl	Ph
example	+**	+	+	–	+
sense	+	–	+	–	+

Table 1: MLU-positioning in dictionaries of Slavic languages.

L = lexicalized units section, Ph = phraseological section, Phl = phraseological list, Pa = paremiological section, Pl = paremiological list, T = terminological section.

* MLUs are treated as headwords of a special type of entries that are not on the same level as single-word entries.

** Graphically separated from other examples by a slash character.

Headword positioning is not used in these dictionaries; also treating MLUs as subheadwords in the Slovak dictionary is limited to specific types of MLUs. The Polish dictionary is somewhat specific. Due to its digital nature it is difficult to tell whether MLUs are treated on microstructural or macrostructural level in this dictionary. This is expected since the boundaries between micro- and macrostructure can be fuzzy in dictionaries designed primarily for use in digital environments in contrast to printed dictionaries (cf. Hausmann & Wiegand 1989, Svensén 1993).

The most frequent placing for MLUs in general monolingual explanatory dictionaries of Slavic Languages are in sections of various types, and in examples. Placing MLUs among examples clearly shows their semantic relation to the senses within which they are presented. It also contributes to the compact data representation of all the different types of language units on the wide border between single-word units and MLUs such as syntactical patterns, units where only

pragmatic information needs to be added for the human user, figurative meanings, etc. Placing MLUs among examples is very typical of print dictionaries, but it does not need to be a common solution for digital dictionaries either due to different media used for publishing or due to the conceptual differences and advances in linguistic and lexicographic theories.

MLU sections in dictionaries are presented either after the last sense (sense-independent) or within a specific sense. While the Slovenian and Croatian dictionary place MLUs in two different types of sections, namely the phraseological and terminological section, the Slovak dictionary uses a four-type section division, and the Russian dictionary uses a one-type section division. The Polish dictionary also places MLUs in two different microstructural sections but the typology of MLUs differs from the typology used in Slovenian and Croatian dictionaries. In three dictionaries MLUs can also be treated at the sense level, namely in SSKJ, SSSJ and TSRYA. One of the main reasons for different solutions is the great diversity of MLU types with many borderline cases and the variety of theoretical approaches.

The most unique approach is used in the Polish dictionary designed for use in a digital environment. The MLUs are presented in two types of sections where the MLUs are listed and, by use of hyperlinking, cross-referenced to a separate MLU-type of entries. Content-wise, the Polish dictionary is the only one of the dictionaries analysed which treats the MLUs as equivalent to single-word lexical units. Beside definitions, exemplification, usage information, and labels, which are all often used also in other dictionaries, other information is also given in a way similar to single-word headwords, such as ontology, lexical relations, a broader representation of formal variants, and sporadically also etymology. Such an approach is straightforward from the current linguistic perspective, since both single- and multi-word lexical units represent basic elements of the lexicon of a language. Beside linguistic reasons, such approaches to dealing with MLUs seem more reasonable for use in electronic dictionaries than in printed dictionaries (all of the other dictionaries analysed were originally print dictionaries), due to the absence of size restrictions and the possibility of the use of hyperlinks.

Based on the analyses of general monolingual explanatory dictionaries of Slavic languages we can conclude that typologically similar MLUs in these dictionaries are treated in various microstructural sections, which is also the case of English dictionaries (see above). This clearly shows that no absolute consensus exist among lexicographers regarding the position in the dictionary entry where the various MLU types are to be treated and what is the preferred typology of MLUs from a lexicographic perspective. It also suggests important implications for the user perspective. The fact that MLUs are presented in various ways across

dictionaries of various languages contributes to diverse user expectations about where to find MLUs in a certain dictionary. The representation of MLUs in a dictionary is usually presented in the front matter, but many users usually do not read such preliminary explanations.

The analyses presented here support other theoretical and practical lexicographical findings that were considered in the process of building the macrostructural and microstructural organization of dictionary entries in the new Slovenian dictionary of standard language (cf. Petric 2018a, Petric 2018b). The solutions for dealing with MLUs in this dictionary are presented in the next section.

2 Multi-word lexical units in the new general monolingual explanatory dictionary of the Slovenian language

The eSSKJ: Dictionary of the Slovenian Standard Language, 3rd Edition⁷ (hereinafter: eSSKJ) is a new medium-sized academic general monolingual explanatory dictionary of the Slovenian standard language (Ledinek 2015; Gliha Komac et al. 2016). Although this is the 3rd edition of the dictionary, it has a completely new conceptual design compared to the previous editions. It is being compiled at the Fran Ramovš Institute of the Slovenian Language of the Scientific Research Centre of the Slovenian Academy of Sciences and Arts. Parts of the dictionary have already been published (Gliha Komac et al. 2017; Gliha Komac et al. 2018, Černivec et al. 2019), and are available on the Fran dictionary portal (www.fran.si).

The eSSKJ is primarily intended for adult native speakers of the Slovenian language. The essential materials of the dictionary are the contemporary (mostly written) corpora of the Slovenian language, namely the Gigafida, KRES, and Nova beseda, while other corpora (KAS, SIWaC, GOS) are used as sources for additional information. The dictionary provides a detailed description of a widespread contemporary Slovenian lexicon, which is based on the interpretation of texts published since 1991, i.e. after the publication of the first edition of the dictionary (SSKJ). Such selection of materials is based on the fact that textual materials from earlier period are not available in digital format. Also, the timeframe coincides with the creation of Slovenian state and changed socio-economic and political circumstances. Although the dictionary is conceived as the dictionary of the standard Slovenian language, it includes a relatively wide set of the most commonly used contemporary Slovenian lexical units (e.g. spoken language,

⁷ The concept of the dictionary (Gliha Komac et al. 2015) in which the principal conceptual features of the dictionary and the envisaged dictionary conventions are clarified, is available on: <http://www.fran.si/novi-sskj>.

dialectal expressions, terminological expressions, etc.) according to their use in standard texts. The lexicon and other linguistic elements included in the dictionary are evaluated and described in relation to the recognized current standard language norm.

In accordance with modern lexicographic practice, eSSKJ and its databases are designed primarily for use in web environments and other digital environments, not only for lexicographic purposes but also for the use of data in linguistic research, development of language technologies, and natural-language processing. The databases the dictionary is based on are written in an internal XML format and structured in a way that they can be converted to any of the standard formats for structuring lexical databases, e.g. TEI. The dictionary databases enable connectivity to a wide variety of other databases. On the basis of new analyses and user inquiry data, the dictionary data can be regularly updated.

2.1 The macrostructure of the eSSKJ dictionary

In order to take into account the Slovenian and Slavic lexicographic tradition and the habits and language practices of Slovenian language-handbook users, as well as the findings of the research on English learners' dictionaries and the bilingual English-Slovenian dictionary (see above), MLUs are not treated at the macrostructural level in eSSKJ, and this is also the case in many dictionaries for other Slavic languages (see above; cf. also Kočová et al. 2014). On the contrary, MLUs are placed into the dictionary microstructure in eSSKJ, except for three types of structures that can also be interpreted as single-word lexemes and are considered MLUs neither in the dictionary nor in this article:

- structures of verbs with reflexive morphemes (e.g. *bati se*; *smejati se*; *zapomniti si*; *znočiti se*),
- loan words written as multi-words according to the orthographic rules of their languages of origin (e.g. *de facto*; *alma mater*; *ad hoc*; *bossa nova*; *homo sapiens*; *urbi et orbi*),
- and orthographically unclear clusters (e.g. *hokus pokus*; *na svidenje*; *kdor koli*).

Despite the fact that in eSSKJ only single-word lexical units are afforded head-word status, eSSKJ places a much greater emphasis on MLUs than the previous general monolingual explanatory dictionary of the Slovenian language. Semantically non-transparent MLUs are treated in two separate microstructural sections of the dictionary (non-phraseological and phraseological section) and are regarded as equivalent to single-word lexical units in the sense that they require as systematic and comprehensive a dictionary description as single-word lexical units. Therefore, eSSKJ affords single-word lexical units as well as semantically

non-transparent MLUs the same types of data, for example: labels, explanations, synonyms, examples, normativity information, grammatical preferences and limitations, syntactic patterns in which MLUs typically occur, etc. The inclusion of a MLU into the dictionary is based on the analysis of its use in contemporary language corpora. Among various types of MLUs, only semantically non-transparent MLUs are explicitly treated as such in eSSKJ, since the primary role of the general monolingual explanatory dictionary is to provide a systematic description of the senses of the core lexical units of the modern Slovenian language. Semantically transparent MLUs are included among collocations and (parts of) examples illustrating use of individual senses and sub-senses of single-word lexical unit since no additional information on these units needs to be provided to the user.

Semantically non-transparent MLUs in eSSKJ are listed in two separate microstructural sections: in a common section for phraseological and paremiological units and in the section for all other (non-phraseological) MLUs. MLUs are placed under all headwords that are the constituents of a MLU, except for function words (*rdeč kot kuhan rak* ‘as red as a lobster’ is therefore included in the entries *rdeč* ‘red’, *kuhan* ‘cooked’ and *rak* ‘crab’, but not the conjunction *kot* ‘as’). This MLU placement is offered to the dictionary user. For editorial reasons, however, the treatment of MLUs in the database is rather different, since they are treated as a type of separate special entries in the dictionary database and cross-referenced to relevant entries. The reason for such a solution is the finding that in some dictionaries (especially in the pre-electronic ones) inconsistencies can be found when treating the same MLU in different places. The separate treatment of MLUs in the database prevents such events.⁸ Upon the exportation of the database for the published version, however, MLUs are included in the entries of their relevant constituents based on the cross-references to the relevant single-word.⁹ It

⁸ Other types of mistakes are still possible, such as the existence of MLU variants as separate entries in the database. However, such mistakes can successfully be found during the editorial and proof-reading process of the single-word entries with MLUs already automatically included upon export.

⁹ The process of cross-referencing is semi-automated by first tokenizing the MLU main form (and its variants, if any), then removing duplicate words (tokens) and finally structuring the information to appropriate XML structure. Each word (token) is given its own structural element, where an attribute containing the word form is added, and the text value (token) is automatically added to be changed to the lemma form by the author (if necessary). Cross-referencing is done using the lemma form as the search string to match the appropriate single-word entry. Target-entry id is used to establish the cross-reference. Upon export, the word form is used to distinguish whether it belongs to the dictionary form of the MLU or only to its variant(s). This information is then presented differently to the dictionary user.

would be therefore possible, based on the existing database, to quickly and easily create a dictionary (or several dictionaries) with a different macrostructural arrangement.

The decision for the treatment of MLUs at the microstructural level in eSSKJ affects the macrostructure of the dictionary as well. Some words in Slovenian language are used predominantly or exclusively as a part of MLUs. For example, the words *deoksiribonukleinski* and *kriplje* are used in MLUs *deoksiribonukleinska kislina* and *na vse kriplje/na vse pretege* only; therefore, no sense definition for single-word use is given. The word *jesihar*, which is used very frequently as part of the phraseme *vpiti/dreti se kot jesihar*, can be used as a single-word lexical unit but is so rare and limited in use that it is not reasonable to include its senses in eSSKJ. In order to be able to show the morphological characteristics (inflection), the pronunciation, and the accentuation¹⁰ of such MLU constituents, they are also given headword status, but only the use of their respective MLUs is described.

There is another argument for presenting MLUs at microstructural level in eSSKJ instead of affording them independent entry status. The eSSKJ is published on the Fran dictionary portal. As of May 2019, the portal consists of 36 dictionaries. On a simple search, all matching results are shown from all the dictionaries. Due to the sometimes large number of search results, the question arises whether it is more efficient to present all MLUs from the dictionary as individual search results (i.e. individual entries) or to join them to the single-word lexical-unit entries in order to make the search result list more compact. For example, the word *buča* ‘pumpkin’ has 19 search results from different dictionaries on the portal. In eSSKJ, it contains 16 MLUs and presenting them as stand-alone entries would increase the number of search results to 35, thus making it difficult for the user to scroll to the search results from any other dictionary. In a stand-alone web dictionary, no such limitation might be needed, cf. above the solution in the Great Dictionary of Polish. The present solution in the Fran portal gives the user a condensed eSSKJ dictionary entry containing core information from the dictionary including a list of MLUs in the entry without further information, as shown in Figure 3. However, the MLUs are clickable and a user can immediately obtain the required information by clicking the MLU and viewing the whole dictionary entry (the ‘Full Entry’ view).

2.2 Placement of MLUs in the microstructure of the eSSKJ dictionary

As we have already mentioned, semantically non-transparent MLUs are placed within two microstructural sections of the dictionary, one being a common section

¹⁰These characteristics are provided for single-word lexical units only in eSSKJ.

for phraseological and paremiological MLUs and the other one being the section for all other (non-phraseological) MLUs. These MLUs are treated separately from all other semantically transparent MLUs. The placement of semantically non-transparent MLUs in two microstructural sections of the dictionary represents a departure from the established Slovenian lexicographic practice, since lexical units that were considered terminological units initially are not treated within a specific microstructural section of eSSKJ, namely the terminological section, as was the case in SSKJ, but in accordance with other non-phraseological MLUs, which is a more common lexicographic practice (Kržišnik 2009: 54). The decision to organize the dictionary data in a different way than in SSKJ was based on the fact that the lexicon included in eSSKJ that was initially part of the terminology of various fields of science has been determinologised (cf. Žagar 2005, Nová 2018) and is thus a part of the general lexicon of the language. The initial terms are described in eSSKJ only if they frequently occur in texts that have a relatively wide circle of addressees. The description of the determinologised lexicon in eSSKJ therefore complies with the principles of lexicographic rather than terminographic treatment. Consequently, there is no reason to place this type of lexical unit in separate microstructural sections of entries, although they are regarded as a somewhat specific segment of the general lexicon.

búča búče *samostalnik ženskega spola* [búča] Full entry eSSKJ

1. kulturna rastlina s plazečim se stebлом in navadno užitnimi plodovi trebušaste oblike;
primerjaj lat. *Cucurbita pepo*

1.1. plod te rastline
1.2. ta plod kot hrana, jed

2. posoda, steklenica trebušaste oblike
2.1. stekleni del svetila take oblike

3. knjižno pogovorno, ekspresivno glava, zlasti človeška

4. ekspresivno neumen, naiven, nespameten človek

STALNE ZVEZE: búča golica, búča hokaido, búča špagetarica, hokaido búča, muškata búča, orjaška búča, presesalna búča, špagetna búča

FRAZEOLOGIJA: *brihtna búča, dobiti jih po buči*, kotaliti búče, kot svinja v búče, *prodajati búče* (komu), *trda búča, trde búče*, To so búče.

ETIMOLOGIJA: = hrv. *bŭča*, prevzeto iz roman. < vulglat. *buttia* 'okrogla posoda za vino, sodček' k *buttis* 'sod' - več ...

Figure 3: The word *buča* 'pumpkin' and its MLUs (in: *Stalne zveze* 'Multi-word lexical units', *Frazeologija* 'Phraseology') among search results

In eSSKJ, three types of MLUs are treated in the section of non-phraseological MLUs: a) non-terminological semantically non-transparent MLUs (e.g. *stara mama*; *kačji pastir*), b) terminologised terms labelled with terminological labels (e.g. iz psihologije *patološki afekt*), and c) the most commonly used idiomatic multi-word proper names (e.g. *Bela hiša*). Regardless of the typology presented, these units are listed in alphabetical order, so that users can find them as easily as possible.

The eSSKJ exhibits the most comprehensive and the most systematic treatment of phraseology in Slovenian general monolingual dictionaries. The dictionary deals with phraseological MLUs of the following types: a) non-sentential phrasemes (e.g. *poskusni zajček*; *stric iz ozadja*; *odpasti kot zrela hruška*; *prodajati bučke (komu)*) and sentential phrasemes (e.g. *srce pade v hlače (komu)*), b) comparative phrasemes (e.g. *rdeč kot kuhan rak*; *težak kot beton*), c) some pragmatic phrasemes (e.g. *gromska strela*; *pojdi se solit*), d) minimal phrasemes (e.g. *popihati jo*; *do amena*) and e) paremiological units (e.g. *Jabolko ne pade daleč od drevesa*).

In addition to the information presented in all other types of single-word and multi-word lexical units (e.g. semantic and pragmatic descriptions, grammatical preferences and limitations, labelling, examples, etc.) the description of MLUs placed in the phraseological section includes the main form of the phraseme (e.g. *prodajati bučke (komu)*), its frequently used variants (e.g. *prodajati buče (komu)*, *prodajati bučnice (komu)*), its optional components, and, if necessary, information on valency. The dictionary also provides a condensed representation of the phraseme, in which (all of) its variants and optional elements are shown (e.g. *prodajati bučke/buče/bučnice (komu)*). Each variant of the phraseme is written in full, which enables users to find them easily. MLUs are listed in alphabetical order within a single-word entry. Paremiological units are placed after all the other phraseological MLUs. The variant forms of the phraseme are sorted according to the frequency of their use, as reflected in modern language corpora.

<p>EF prodajati bučke (<i>komu</i>) <i>prodajati buče (komu)</i> <i>prodajati bučnice (komu)</i> STRNJENI PRIKAZ: prodajati bučke/buče/bučnice (<i>komu</i>) govoriti, pisati neumnosti, besedne nesmisle; zavajati koga z govornjem, pisanjem neresnic, neumnosti Naši politiki nam prodajajo bučke, da se ne poznajo med sabo in da niso vpleteni v nič. Daj, pojdi raje na igrišče in ne prodajaj buč, ker nismo vsi ljudje neumni in naivni. Tu pa je problem, ki ga mora rešiti vlada, a ga noče, raje nam prodaja bučnice o (ne)konkurenčnosti.</p>

Figure 4: Description of phraseological MLUs in eSSKJ.

3 Conclusion

A variety of MLU types present lexicographers with several questions that are to be dealt with before the dictionary-compilation process and further elaborated in the later stages of compilation. The analyses of treatment and placing of MLUs in Slavic dictionaries shows the diverse solutions regarding the positioning and representation of MLUs. The differences arise from different theoretical backgrounds, MLU typologies, the medium in which a dictionary was first published, and the user perspective. The shift from paper to electronic dictionaries and new theoretical findings enabled dictionary authors to treat MLUs similarly (or equally) to single-word lexemes by describing them with nearly the same types of linguistic information while at the same time making them relatively independent units with regard to their macro- and microstructural position within the dictionary structure. The different structural solutions between digital dictionaries can also be a consequence of the context in which the dictionary is available, i.e. whether the dictionary on the web is a part of a larger portal or not. For all these reasons, we can also expect different lexicographic solutions for the description of MLUs in the future.

References

- Ahačič, Kozma, Ledinek, Nina, Perdih, Andrej. 2015. Fran: The Next Generation Slovenian Dictionary Portal. In K. Gajdošová, A. Žáková (eds.) *Proceedings of the 8th International Natural Language Processing, Corpus Linguistics, Lexicography Conference*, 9-16. Accessed at: http://korpus.sk/~slovko/2015/Proceedings_Slovko_2015.pdf [31/03/2018].
- Anić, Vladimir. 2004. *Veliki rječnik hrvatskoga jezika*. Zagreb: Novi Liber.
- Atkins, Sue B.T., Rundell, Michael. 2008. *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Černivec, Manca et al. 2019. *Slovar slovenskega knjižnega jezika 2018*. Ljubljana: ZRC SAZU. Accessed at: <https://isjfr.zrc-sazu.si/sites/default/files/sskj-2018.pdf>. [31/03/2019].
- Fran, slovarji Inštituta za slovenski jezik Frana Ramovša ZRC SAZU, 2014–, version 5.0. Accessed at: www.fran.si [31/03/2018].
- Gliha Komac, Nataša et al. 2015. *Koncept novega razlagalnega slovarja slovenskega knjižnega jezika*. Ljubljana: Inštitut za slovenski jezik Frana Ramovša ZRC SAZU. Accessed at: <http://www.fran.si/novi-sskj> [31/03/2018].
- Gliha Komac, Nataša et al. 2016. Novi slovar slovenskega knjižnega jezika – predstavitev temeljnih konceptualnih izhodišč. In Marušič, Franc, Mišmaš, Petra, Žaucer, Rok (eds.) *Zbornik prispevkov s simpozija 2015*. Nova Gorica: Založba Univerze, 16-33. Accessed at: http://www.ung.si/media/storage/cms/attachments/2016/10/21/13/45/01/Zbornik-%C5%A0D9_okt.16_splet.pdf [31/03/2018].

- Gliha Komac, Nataša et al. 2017. *Slovar slovenskega knjižnega jezika 2016*. Ljubljana: ZRC SAZU. Accessed at: https://isjfr.zrc-sazu.si/sites/default/files/sskj2016-20-6-2017_0.pdf [29/03/2018].
- Gliha Komac, Nataša et al. 2018. *Slovar slovenskega knjižnega jezika 2017*. Ljubljana: Založba ZRC, ZRC SAZU. Accessed at: https://isjfr.zrc-sazu.si/sites/default/files/sskj_2017.pdf [29/03/2018].
- Hausmann, Franz Josef, Wiegand, Herbert Ernst, 1989. Component Parts and Structures of General Monolingual Dictionaries: A Survey. V: HAUSMANN, Franz Josef idr. (ur.): *Wörterbücher / Dictionaries / Dictionnaires. Ein internationales Handbuch zur Lexikographie / An International Encyclopedia of Lexicography / Encyclopédie internationale de lexicographie. Handbücher zur Sprach und Kommunikationswissenschaft / Handbooks of linguistics and communication science / Manuels de linguistique et des sciences de communication, Band 5.1*. Berlin, New York: W. de Gruyter. 328–360.
- Kochová, Pavla, Opavská Zdeňka, Holcová Habrová, Martina. 2014. At the Beginning of a Compilation of a New Monolingual Dictionary of Czech (A Report on a New Lexicographic Project). Abel, Andrea et al. (eds.): *Proceedings of the XVI EURALEX International Congress: The User in Focus*. Bolzano/Bozen: Institute for Specialised Communication and Multilingualism. 1145-1151.
- Kržišnik, Erika. 2009. Kam in kako s frazeologijo v novem slovarju. In A. Perdih (ed.) *Strokovni posvet o novem slovarju slovenskega jezika, 23. in 24. oktober 2008*. Ljubljana: Založba ZRC, ZRC SAZU. 53-56.
- Ledinek, Nina. 2016. Novi Slovar slovenskega knjižnega jezika: temeljna konceptualna izhodišča. In: Ristić, Stana et al. (eds.) *Leksikologija i leksikografija u svetlu savremenih pristupa: zbornik naučnih radova*. Beograd: Institut za srpski jezik SANU. 119-130.
- Legan Ravnikar, Andreja. 2017. K problematiki vpliva stičnega jezika – nemščine na semantične spremembe in stilno vrednost najstarejše slovenske knjižne leksike (16. stoletje). In *Slovenski jezik / Slovene Linguistic Studies* 11: 35-53.
- Lew, Robert. 2012. How can we make electronic dictionaries more effective?. In S. Granger, M. Paquot (eds.) *Electronic Lexicography*. Oxford: Oxford University Press. 343-363.
- Nová, Jana. 2018. Terms Embraced by the General Public: How to Cope with Determinologization in the Dictionary?. Čibej, Jaka et al. (eds.): *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts*. Ljubljana: Znanstvena založba Filozofske fakultete Univerze v Ljubljani / Ljubljana University Press, Faculty of Arts. 387-398.
- Oppentocht, Lineke, Schutz, Rik. 2003. Developments in electronic dictionary design. In van Sterkenburg, Piet (ed.) *A practical guide to lexicography, Terminology and Lexicography*. Amsterdam: John Benjamins. 215-227.
- Perdih, Andrej. 2016. *Splošni razlagalni slovarji slovanskih jezikov*. Založba ZRC, ZRC SAZU, Ljubljana.
- Perdih, Andrej, Snoj, Marko. 2015. SSKJ2. In *Slavia Centralis* 8/1: 5-15.
- Petric, Špela. 2018a. *Slovenski večbesedni leksemi z vidika slovaropisja*. Založba ZRC, ZRC SAZU, Ljubljana.
- Petric, Špela. 2018b. Slovaropisni vidik samostalniških nefrazeoloških večbesednih leksemov. In *Philological studies* 16/1: 214-229.
- Shvedova, N. Yu. 2008. *Tolkovyy slovar' russkogo jazyka s vključenjem svedenij o proiszhodenii slov*. Moskva: Rossijskaja akademija nauk.

- Silvester, Marta. 1978. *Priročnik za tehnično stran redakcije gesel*. Ljubljana: Inštitut za slovenski jezik SAZU. Leksikološka sekcija. Tipkopis.
- Slovar slovenskega knjižnega jezika, druga, dopolnjena in deloma prenovljena izdaja. Accessed at: www.fran.si [29/03/2018].
- Slovník súčasného slovenského jazyka*. Bratislava: VEDA. I. (2006): A–G. II. (2011): H–L. (2015): III. (M–N).
- Slovar slovenskega knjižnega jezika*, I–V (1970–1991). Ljubljana: SAZU, DZS.
- Snoj, Jerica. 2019. *Leksikalna sinonimija v sinonimnem slovarju slovenskega jezika*. Ljubljana: Založba ZRC, ZRC SAZU.
- Svensén, Bo. 1997. *Practical Lexicography. Principles and Methods of Dictionary Making*. Oxford/New York: Oxford University Press.
- Vrbinc, Alenka. 2011. Macrostructural Treatment of Multi-word Lexical Items. In *ELOPE: English language overseas perspectives and enquiries* 8: 51–61.
- Vrbinc, Alenka, Vrbinc, Marjeta. 2004. Language Learners and Their Use of Dictionaries: The Case of Slovenia. In *Erfurt Electronic Studies in English*, 3. Accessed at: http://webdoc.gwdg.de/edoc/ia/eese/artic24/marjeta/3_2004.html [29/03/2018].
- Wielki słownik języka polskiego PAN*. Accessed at: <http://www.wsjp.pl> [29/03/2018].
- Žagar, Mojca. 2005. Determinologizacija (na primeru terminologije fizike). In *Jezik in slovnstvo* 50/2: 35–48.
- Żmigrodzki, Piotr. 2014. Polish Academy of Sciences Great Dictionary of Polish [Wielki słownik języka polskiego PAN]. In *Slovenščina 2.0* 2 (2): 37–52.
- Żmigrodzki, Piotr. 2018. Geneza słownika i historia projektu. In P. Żmigrodzki et al. (eds.) *Wielki słownik języka polskiego PAN. Geneza, koncepcja, zasady opracowania*. Kraków: Instytut Języka Polskiego PAN. 9–16.
- Žugić, Radmila V. 2016. Leksikografska interpretacija višočlanih leksičkih jedinica u rečnicima prizrensko-timočkih dijalekata i pitanje njihovog statusa kao posebnih odrednica. In *Jezikoslovní zapiski* 22/1: 153–169.

Received April 2019, accepted June 2019.

Prispelo aprila 2019, sprejeto junija 2019.

Multi-word Lexical Units in General Monolingual Explanatory Dictionaries of Slavic languages

In the article, the typology and the macro- and microstructure positioning of multi-word lexical units in general monolingual explanatory dictionaries of five Slavic languages (Slovenian, Croatian, Slovak, Polish and Russian) are analysed. The research showed that MLUs in these dictionaries are most commonly treated on the microstructural level (with the exception of the newest general monolingual explanatory dictionary of Polish language), where typologically comparable or similar MLUs are treated in various microstructural sections, most commonly among dictionary examples and in various types of MLU sections. The differences in the treatment of MLUs can arise also from the specifics of the medium in which a dictionary was first published. The dictionaries built primarily for web and other digital environments are based on structured machine readable databases, therefore the MLUs are regarded as equivalent to single-word lexical units in the sense that they require as systematic and comprehensive a dictionary description as single-word lexical units. Consequently, the same types of data are normally afforded to these units. At the same time, such a shift in the treatment of MLUs can also be attributed to the development of lexicology, lexicography and meta-lexicography. In the newest dictionaries, the treatment of MLUs is influenced also by the research on user perspective and the possibility to incorporate a dictionary into language portals, while the comprehensive treatment of MLUs is motivated also by the potential of the dictionary data for the linguistic research, and the development of language technologies and natural language processing. With regard to the findings about the typology and positioning of MLUs in these dictionaries, the article also focuses on the treatment of MLUs in eSSKJ: The Dictionary of the Slovenian Standard Language, 3rd edition, the new monolingual general explanatory dictionary of the Slovenian language. In this dictionary, MLUs are treated in a similar way to single-word lexical units and are given relative autonomy in the dictionary structure.

Večbesedne leksikalne enote v splošnih enojezičnih razlagalnih slovarjih slovanskih jezikov¹¹

V članku obravnavamo tipologijo večbesednih leksikalnih enot v enojezičnih splošnih razlagalnih slovarjih petih slovanskih jezikov (slovenskem, hrvaškem, slovaškem, poljskem in ruskem) in njihovo umestitev v slovarsko makro- oz. mikrostrukturo. Analiza je pokazala, da so večbesedne leksikalne enote v teh slovarjih najpogosteje obravnavane na mikrostrukturni ravni, izjema je le naj-novejši poljski enojezični splošni razlagalni slovar, tipološko enake ali podobne večbesedne leksikalne enote pa so opisane v različnih mikrostrukturnih razdelkih slovarjev, najpogosteje med slovarskimi zgledi in v specializiranih mikrostrukturnih razdelkih. Razlike v opisu večbesednih leksikalnih enot se kažejo tudi z vidika medija, za katerega je slovar primarno oblikovan – v novejših slovarjih, namenjenih rabi v spletnem okolju in drugih digitalnih okoljih, ki temeljijo na strukturiranih računalniško berljivih podatkovnih bazah, so večbesedne leksikalne enote razumljene kot enobesednim leksikalnim enotam enakovredne enote, ki potrebujejo sistematičen in celovit slovarski opis. Pri njihovem opisu so zato načeloma navedeni vsi tisti slovarski podatki, ki jih dodajamo tudi enobesednim leksikalnim enotam. To spremembo je obenem mogoče pripisati tudi razvoju leksikologije, leksikografije in metaleksikografije. Seveda na obravnavo večbesednih leksikalnih enot zlasti v najnovejših slovarjih vplivajo tudi uporabniške zahteve in možnost vključitve slovarskih virov v večje jezikovne portale, celovita obravnava teh enot pa je smiselna tudi z vidika uporabe slovarskih podatkov za jezikoslovne raziskave in za razvoj jezikovnih tehnologij oz. pri procesiranju naravnih jezikov. Članek se osredotoča tudi na večbesedne leksikalne enote v eSSKJ: tretji izdaji Slovarja slovenskega knjižnega jezika, najnovejšem enojezičnem razlagalnem slovarju slovenščine, v katerem so te enote obravnavane kot relativno samostojne enote, opisane na podoben način kot enobesedne leksikalne enote.

¹¹ This article has been supported by ARRS (program P6-0038).